TABLE S1 Read numbers of *amoA/pmoA* genes of AOA, AOB, comammox and MOB in various environmental samples from metagenomic or metatranscriptomic datasets in the NCBI SRA database. (Excel file format)

TABLE S2 Read numbers of *amoA/pmoA* genes of AOA, AOB, comammox and MOB in marine samples from metagenomic or metatranscriptomic datasets in the CAMERA database. (Excel file format)

TABLE S3 Read numbers of comammox and AOB in 18 different environmental samples (amplification failed for the other three of the 21 samples) analyzed via partial nested PCR amplification of the comammox-specific *amoA* gene.

Sample type	No. of reads	clade A		clade B	AOB	Others
		clade A.1	clade A.2			
Soil						
Dry land soil	57722	129	3626	125	51985	1857
Coal mine soil	17448	100	2334	10	13303	1701
Landfill soil	99747	5478	2111	182	91829	147
Paddy soil	147776	16311	11990	23340	96103	32
Plateau soil	20440	4715	143	13102	2477	3
Forest soil	41931	86	1368	38149	2325	3
Sediment						
Tidal flat sediment	6853	5193	1126	10	522	2
Saltmarsh sediment	20498	10483	6982	683	2319	31
Coastal sediment	36591	18909	8926	39	8714	3
River sediment	20132	7933	2930	7657	1609	3
Lake sediment	198760	175734	12033	2404	8526	63
Sludge						
Activated sludge	17130	372	43	13	16699	3
Anaerobic sludge	19194	13027	896	14	5247	10
Water						
Tap water	27302	3302	23663	27	305	5
Coastal water	29191	25144	3911	63	67	6
River water	22134	11156	8466	182	2324	6
Lake water	4958	4252	341	7	47	311
Leaf						
Leaf surface	28839	13899	7211	718	6991	20

TABLE S4 The 30 major OTUs (≥ 0.7%) of the comammox *amoA* gene with 90% similarity cutoff for each environment sample. Please see Figure 3 for the phylogeny of these major OTUs. (Excel file format)

TABLE S5 Amino acid similarity matrix of different comammox OTUs identified with a 90% similarity cutoff. (Excel file format)

TABLE S6 Diversity indices with 97% or 90% similarity cutoffs for comammox, AOB and AOA *amoA* gene sequences in each of the 12 environmental samples after rarifying read numbers to 1900 for each sample. (Excel file format)

TABLE S7 Characteristics of various environmental samples analyzed in this study

Sample ID	Sample name	Sample type	Location	Longitude and latitude	Sample description	Sampling time	pH	NO_3^- (μmol/g)	NH_3 (μmol/g)
1	Dry land soil	soil	Chongming Island, Shanghai, China (Yangtze River estuary)	121°36' E, 31°36' N	0-5 cm of surface soil planted with wheat	Oct. 2013	- [a]	-	-
2	Coal mine soil	soil	Shanxi Province, China (the Loess Plateau)	109°35' E, 38°26' N	0-5 cm of surface soil covered by grass near the operating fan outlet	May 2012	7.9	0.10	0.00
3	Landfill soil	soil	Disused municipal solid waste landfill, Shanghai, China	121°52' E, 31° 02' N	0-5 cm of surface soil planted with shrubs	Apr. 2011	7.1	4.22	0.50
4	Paddy soil	soil	Chongming Island, Shanghai, China	121°36' E, 31°36' N	0-5 cm of surface soil planted with rice	Oct. 2011	7.64	0.70	1.16
5	Glacier soil	soil	Xinjiang Province, China	86°50' E, 43°07' N	0-5 cm of surface soil covered by ice	Aug. 2009	-	-	-
6	Plateau soil	soil	Tibet, China (high elevation)	94°25' E, 29°38' N	0-5 cm of surface soil planted with grass	Aug. 2009	8.7	0.09	0.13
7	Forest soil	soil	Forest Park, Shanghai, China	121°33' E, 31°19' N	0-5 cm of surface soil planted with deciduous trees	Oct. 2013	8.1	2.08	0.39
9	Tidal flat sediment	sediment	Chongming Island, Shanghai, China	121°59' E, 31°29' N	0-5 cm of surface sediment (no plant)	Oct. 2013	7.32	0.09	0.05
10	Saltmarsh sediment	sediment	Chongming Island, Shanghai, China	121°57' E, 31°31' N	0-5 cm of surface sediment planted with reed at intertidal region	Oct. 2013	7.6	0.02	0.21
11	Coastal sediment	sediment	East China Sea, China (Near Shanghai)	122°07' E, 31° 02′ N	0-10 cm of surface sediment at 20 m of water depth	Jul. 2011	7.45	0.03	0.07

12	River sediment	sediment	Lower reach of Yangtze River, Shanghai, China	121°45' E, 31°28' N	0-5 cm of surface sediment at 0.5 m of water depth	Oct. 2013	7.6	3.88	0.00
13	Lake sediment	sediment	Ping Lake, Zhejiang Province, China (Middle-Lower Yangtze Plain)	120°01' E, 30°41' N	0-5 cm of surface sediment at 2 m of water depth	Aug. 2012	7.5	9.66	1.37
14	Wetland sediment	sediment	Songduo wetland, Tibet, China (Tibetan Plateau)	100°35' E, 32°15' N	0-5 cm of surface sediment of plateau wetland covered by grass	Aug. 2009	7.46	20.78	5.65
15	Hot spring sediment	sediment	Riduo hotspring, Tibet, China (Tibetan Plateau)	99°40' E, 35°13' N	0-5 cm of surface sediment at 60°C	Aug. 2009	7.84	0.35	0.00
16	Activated sludge	sludge	Hongkou Qu, Shanghai, China	121°29' E, 31°17' N	Aeration tank of a sewage treatment plant	Jun. 2009	7.52	0.24	16.06
17	Anaerobic sludge	sludge	Brewage wastewater treatment plant, Shanghai, China	121°33' E, 31° 04' N	Recycling sludge from anaerobic digestion plant	Jun. 2009	7.6	1.10	13.87
18	Tap water	water (filter)	Fudan University, Shanghai, China	121° 31′ E, 31° 18′ N	1 L water filtered through a 0.22-μm nitrocellulose filter	Sep. 2015	7.5	0.00	0.00
19	Coastal water	water (filter)	Coast of East China Sea, China (near Zhejiang Province)	122° 13′ E, 28° 45′ N	1 L water filtered through a 0.22-μm nitrocellulose filter	Jul. 2011	-	-	-
20	River water	water (filter)	Lower reach of Yangtze River, Shanghai, China	121°45' E, 31°28' N	0.5 L surface water filtered through a 0.22-μm nitrocellulose filter	Oct. 2013	-	-	-
21	Lake water	water (filter)	Ping Lake, Zhejiang Province, China (Middle-Lower Yangtze Plain)	120°01' E, 30°41' N	1 L of 0-5 cm surface water filtered through a 0.22-μm nitrocellulose filter	Aug. 2012	-	-	-
22	Leaf surface	water (filter)	Pond in Fudan University, Shanghai, China	121°29' E, 31°18' N	1 L sterile water shaken with 3 pieces of lotus leaves (~ 0.3 m^2 total surface)	May, 2011	-	-	-

and filtered through a 0.22-μm nitrocellulose filter

[a] -: not determined.

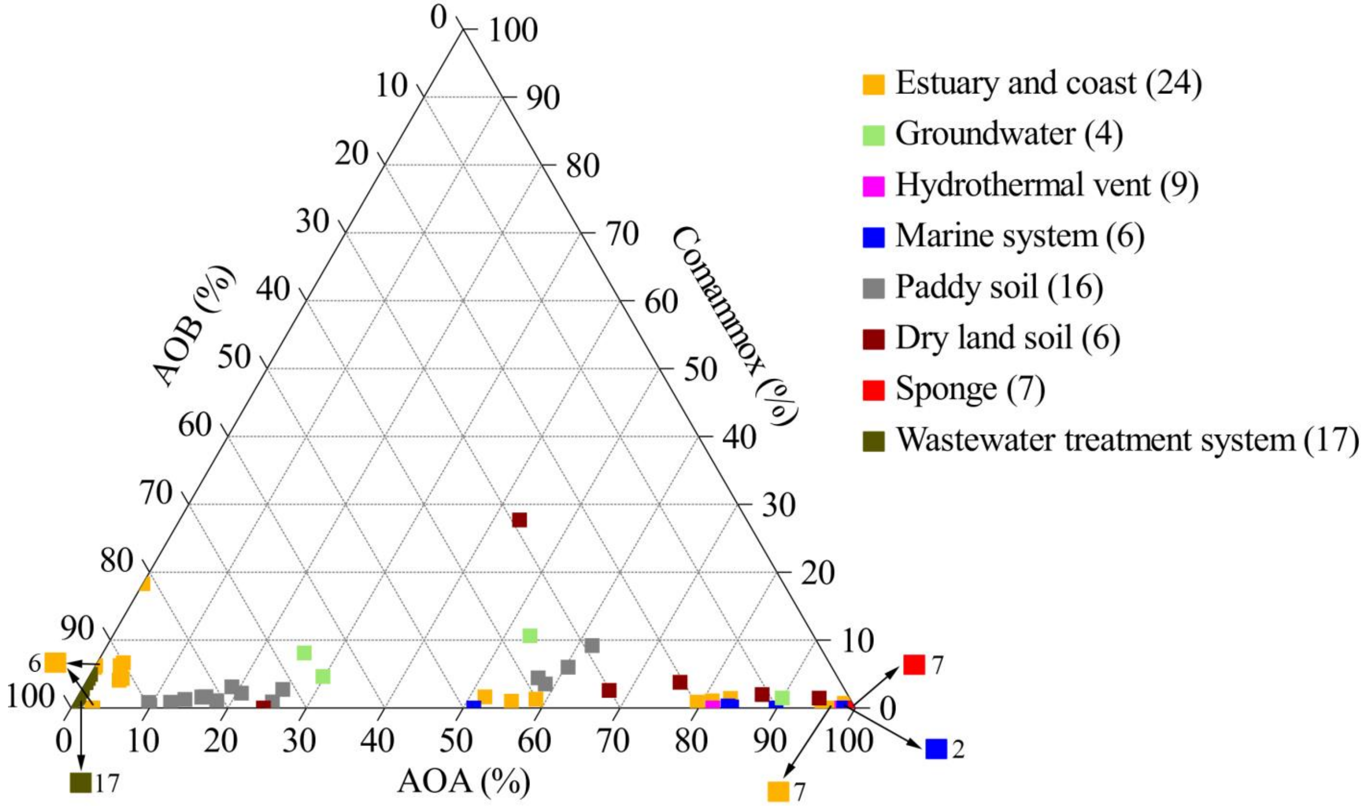

FIG S1 A ternary plot of the relative abundances of comammox, AOB, and AOA in 90 metatranscriptomic datasets of different environmental samples from the NCBI SRA database with *amoA* gene sequence read numbers of at least 30. The numbers directed by arrows in the figure indicate the numbers of overlapping specific color squares. The numbers in parentheses denote the numbers of datasets used in this analysis for specific environmental types. For detailed information on each of the datasets, please see Table S1.

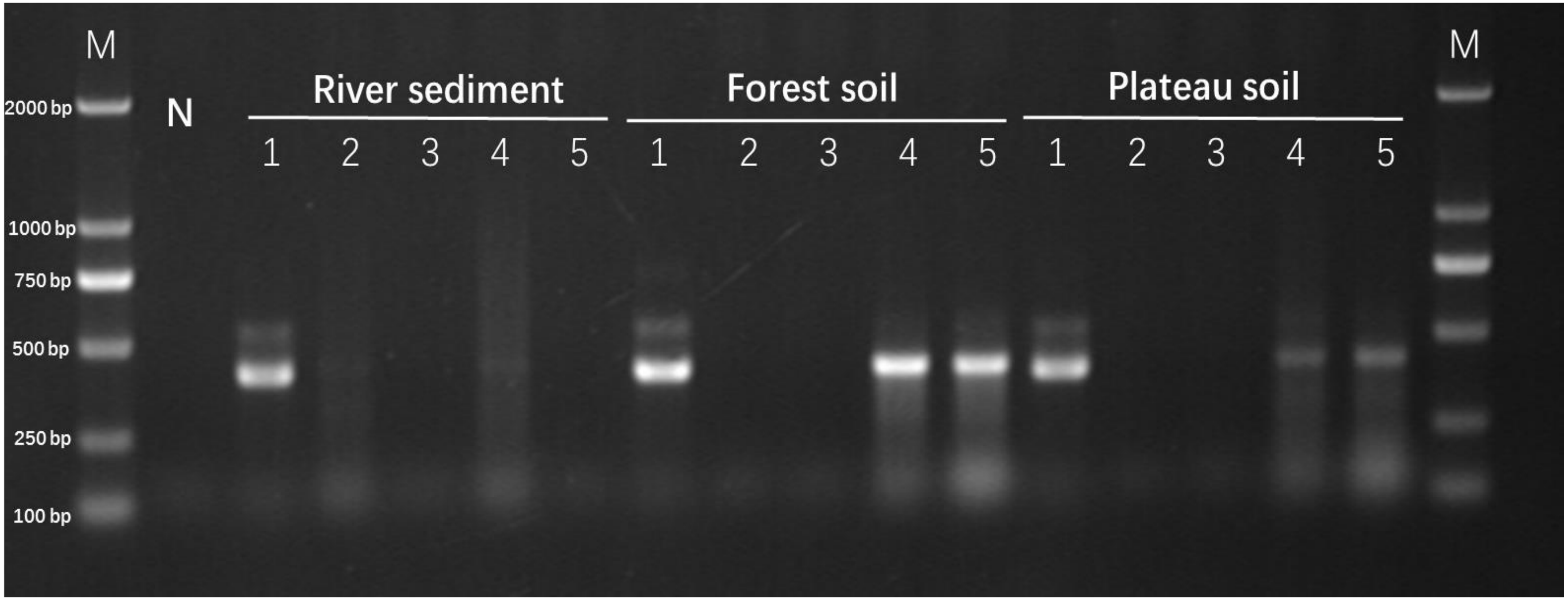

FIG S2 Gel image of different PCR products from river sediment, forest soil, and plateau soil samples. Lanes M and N indicate the marker and negative control for partial nested PCR, respectively, and lanes 1-5 indicate the PCR product using different methods with the different samples mentioned above. Lane 1: partial nested PCR; Lane 2: PCR with degenerate primers for clade A; Lane 3: PCR with mixed primers for clade B; Lane 4: PCR with degenerate primers for clade B; Lane 5: PCR with mixed primers for clade B. The PCR primers and amplification conditions for Lane 1 are described in the Materials and methods section, and those for Lanes 2-5 are described in Pjevac *et al.* (Pjevac et al., Front Microbiol 8:e1508, 2017).
).

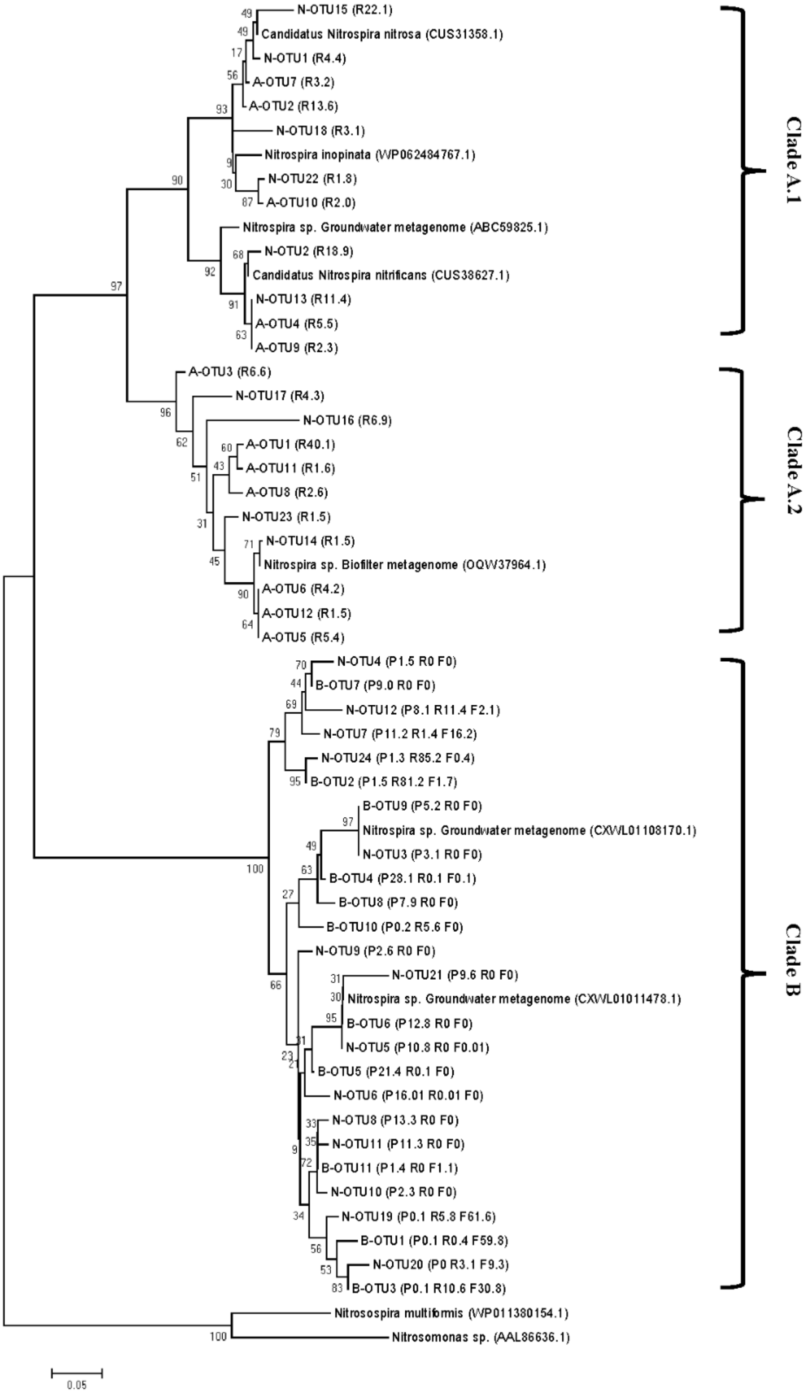

FIG S3 Neighbor-joining phylogenetic tree for comammox AmoA retrieved by partial nested PCR developed in this study and *Nitrospira amoA*-specific degenerate primers developed by Pjevac *et al.* (Pjevac et al., Front Microbiol 8:e1508, 2017). Characters “A”, “B” and “N” in the OTU names indicate the degenerate primers for clade A, degenerate primers for clade B and primers for partial nested PCR, respectively. Characters “P”, “R” and “F” in the parentheses indicate plateau soil, river sediment and forest soil, respectively. The number after the sample symbol character indicates the percentage of the sequences in this OTU of this sample compared to the total sequence number retrieved using the same PCR primers for this sample (from all OTUs).

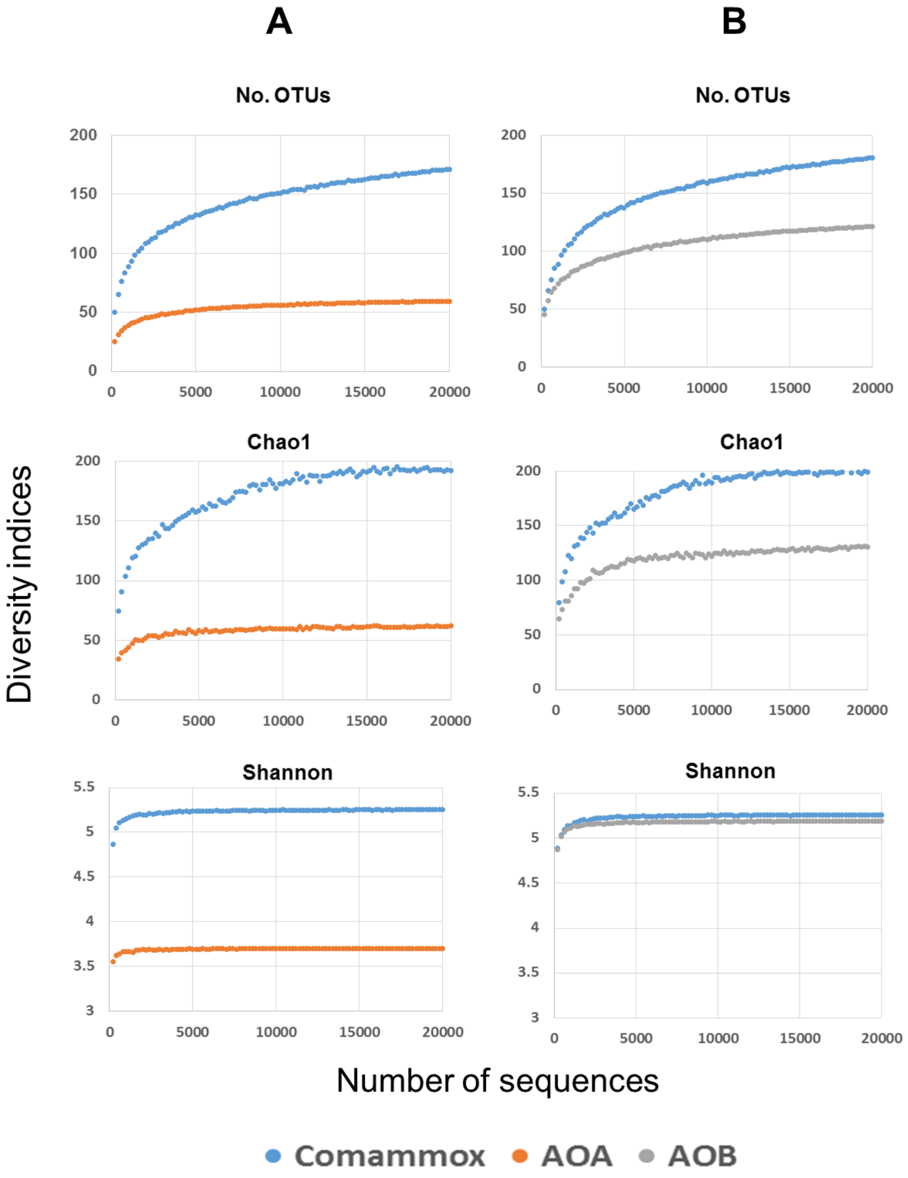

FIG S4 Rarefaction curves of the number of OTUs, the Chao1 estimator and the Shannon index using OTU similarity thresholds of 90% with same region of the *amoA* gene for comparisons of comammox and AOB (A) and comammox and AOA (B) based on a combination of 1900 randomly selected reads from each of the 12 different environmental samples.

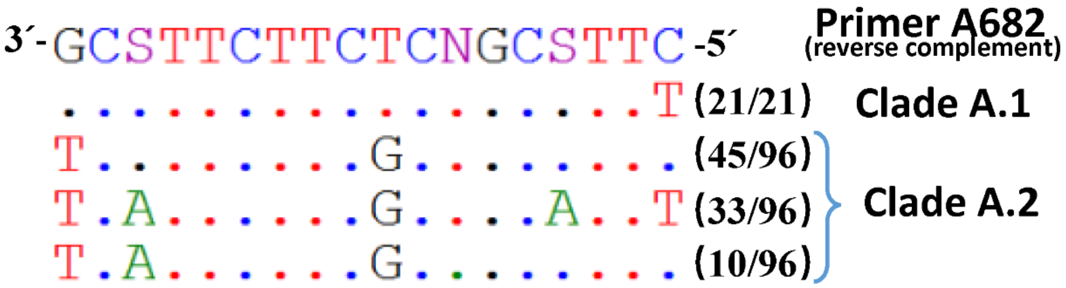

FIG S5 Evaluation of the coverage of primer A682 for comammox bacteria. The sequences were obtained from seven clone libraries of comammox *amoAB* genes amplified using the THDP-PCR method from tidal flat sediment, rice soil, forest soil, leaf surface, tap water, sea water, and activated sludge samples. The data in parentheses indicate the frequency of each sequence in the total sequence sets of clade A.1 and clade A.2.